Week 9

The Gamma 2 Norm and Its Variants

This week, we will study a lower bound technique for quantum communication complexity which is analogous to approximate degree. Actually, in communication complexity, we will have two closely related measures: approximate logrank, and (log of) approximate gamma 2 norm. We will also look at analogues for discrepancy (PP degree). Before we get to that, we start with a few general comments about lifting theorems.

9.1 More on lifting theorems

As mentioned last week, lifting theorems take the form

$$\forall f \quad M^{\mathrm{CC}}(f \circ G) = \Omega(M(f)),$$

or sometimes the form

$$\forall f \quad M^{\mathrm{CC}}(f \circ G) = \Omega(M(f)M^{\mathrm{CC}}(G)),$$

where f is a query function and G is a communication "gadget". Here $M(f)$ is a query measure of f, and M^{CC} is an analogous communication measure.

In order to understand how lifting theorems work, let's start by reviewing some common communication complexity functions. We've already seen the equality function EQ_n, whose communication matrix is the $2^n \times 2^n$ identity matrix. Another common function is called the *inner product* function: here Alice and Bob get strings in $\{0,1\}^n$, and their goal is to compute the inner product of the strings modulo 2. That is to say, if Alice holds x and Bob holds y, they must output $\mathrm{IP}(x,y) = \textsc{Parity}(x_1y_1, x_2y_2, \ldots, x_ny_n)$. Note that the product x_1y_1 is the product in the $\{0,1\}$-basis, meaning the AND function; hence we can write

$$\mathrm{IP}_n(x,y) = \textsc{Parity}_n(x_1\mathrm{AND}y_1, x_2\mathrm{AND}y_2, \ldots, x_n\mathrm{AND}y_n).$$

Note that $\textsc{Parity}_n \colon \{0,1\}^n \to \{0,1\}$ is a query function, while $\mathrm{IP}_n \colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and $\mathrm{AND} \colon \{0,1\} \times \{0,1\} \to \{0,1\}$ are communication functions. Hence, if we take $f = \textsc{Parity}_n$ and $G = \mathrm{AND}$, we can write $\mathrm{IP}_n = f \circ G$. In other words, inner product can be viewed as a *lifted function*: the corresponding query function is Parity, and the lifting gadget is the 2-bit AND function.

Another commonly-encountered function in communication complexity is called *disjointness*. For this function, Alice and Bob each get a string in $\{0,1\}^n$, which they interpret as their schedules over n time slots: if $x_i = 1$, it means Alice is free at the time slot i, and if $x_i = 0$, it means Alice is busy; similarly, if $y_i = 1$, it means Bob is free at time i, and if $y_i = 0$, Bob is busy. The goal

of Alice and Bob is to figure out whether their availabilities are disjoint: whether it is possible for them to meet. In other words, they wish to compute

$$\mathrm{DISJ}_n(x,y) = \mathrm{OR}_n(x_1\mathrm{AND}y_1, x_2\mathrm{AND}y_2, \ldots, x_n\mathrm{AND}y_n).$$

Note that disjointness is once again a lifted function: the corresponding query function is OR_n, and the gadget is the 2-bit AND function.

Now, if we had a lifting function with the AND gadget for some complexity measure – say randomized algorithms – then we could lower bound both IP and DISJ at once: we could write $\mathrm{R}^{\mathrm{CC}}(f \circ \mathrm{AND}) = \Omega(\mathrm{R}(f))$, and this will give both $\mathrm{R}^{\mathrm{CC}}(\mathrm{IP}) = \Omega(n)$ and $\mathrm{R}^{\mathrm{CC}}(\mathrm{DISJ}) = \Omega(n)$.

Unfortunately, a lifting theorem with the AND gadget is not possible: consider the query function $f = \mathrm{AND}_n$. Then $f \circ \mathrm{AND}$ is the communication function where Alice and Bob each get a string in $\{0,1\}^n$, and they must compute the AND of all $2n$ bits they received. This can be done using $O(1)$ communication: Alice sends $\mathrm{AND}_n(x)$ to Bob, and Bob computes $\mathrm{AND}_n(y) \wedge \mathrm{AND}_n(x)$. Hence $\mathrm{R}^{\mathrm{CC}}(\mathrm{AND}_n \circ \mathrm{AND}) = O(1)$, but $\mathrm{R}(\mathrm{AND}_n) = \Omega(n)$, so we cannot have $\mathrm{R}^{\mathrm{CC}}(f \circ \mathrm{AND}) = \Omega(\mathrm{R}(f))$ for all f.

While lifting theorems using 2-bit gadgets like AND, OR, and XOR are impossible, it is sometimes possible to lift using slightly more complex gadgets. Interestingly, this often suffices. For example, consider the gadget $G = \mathrm{DISJ}_2 \colon \{0,1\}^2 \times \{0,1\}^2 \to \{0,1\}$. Note that $\mathrm{DISJ}_n = \mathrm{OR}_{n/2} \circ \mathrm{DISJ}_2$; hence if we had a lifting theorem with $G = \mathrm{DISJ}_2$, then we would have a lower bound on DISJ_n. Similarly, if we had a lifting theorem with $G = \mathrm{IP}_2$, then we would have a lower bound on IP_n.

Interestingly, lifting with a constant-sized inner product gadget and lifting with a constant-sized disjointness gadget are equivalent theorems. This is because for each k, DISJ_{2^k} contains a copy of IP_k as a sub-function (submatrix of the communication matrix), and similarly IP_{2^k} contains a copy of DISJ_k. Hence, if we knew that $\mathrm{R}^{\mathrm{CC}}(f \circ \mathrm{IP}_k) = \Omega(\mathrm{R}(f))$ for all f, then we would also know that $\mathrm{R}^{\mathrm{CC}}(f \circ \mathrm{DISJ}_{2^k}) = \Omega(\mathrm{R}(f))$ for all f, and the latter already gives $\mathrm{R}^{\mathrm{CC}}(\mathrm{DISJ}_n) = \Omega(\mathrm{R}(\mathrm{OR}_{n/2^k})) = \Omega(n)$ when k is constant. Similarly, if we had a lifting theorem with a constant-size disjointness gadget, then we would have a lifting theorem with a constant-size inner product gadget. In fact, both disjointness on 2^k bits and inner product on 2^k bits contain all communication functions on k bits as sub-functions; hence lifting with *any* constant size gadget suffices to lift with both a constant-size disjointness gadget and a constant-size inner product gadget.

Lemma 9.1. *Let $G \colon \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$. Then DISJ_{2^k} contains a copy of G as a sub-function (after removing duplicate rows and columns from G).*

Proof. Note that we can always remove duplicate rows and columns without affecting the communication complexity, since if G has duplicate rows, it means two different inputs x and x' satisfy $G(x,y) = G(x',y)$ for all y, and hence can be treated identically for the purposes of computing G.

Now, let $\phi(x)$ be the string of length 2^k which is all 0 except for a single 1 at the index corresponding to x; that is, $\phi(x)_x = 1$. Let $\psi(y)$ be the string of length 2^k which has $\psi(y)_x = 1$ for all $x \in \{0,1\}^k$ such that $G(x,y) = 1$, and $\psi(y)_x = 0$ for all $x \in \{0,1\}^k$ such that $G(x,y) = 0$. Then $\mathrm{DISJ}_{2^k}(\phi(x), \psi(y)) = 1$ if and only if $G(x,y) = 1$. Moreover, x uniquely determines $\phi(x)$ and vice versa, and y uniquely determines $\psi(y)$ and vice versa. Hence $\phi(x)$ and $\psi(y)$ are just renamings of x and y respectively, so the communication function DISJ_{2^k} restricted to $\{\phi(x) : x \in \{0,1\}^k\} \times \{\psi(y) : y \in \{0,1\}^k\}$ has the same communication matrix as G. □

A similar thing can be shown for inner product, and for many other "sufficiently rich" families of functions. Hence a lifting theorem with a constant-sized gadget is the holy grail of lifting theorems: one you have it, then you effectively also have it with any gadget of your choice, and in particular

you can show disjointness and inner product lower bounds in the model you lifted (assuming you have query lower bounds on OR_n and $\textsc{Parity}_n$ for the model you lifted).

For many computational models, we do not have a lifting theorem with a constant-sized gadget; it remains an open problem. This includes important models such as D^{CC}, R^{CC}, and Q^{CC}. However, we can sometimes show lifting theorems with a gadget G whose size grows with the input size n. A common form of such a theorem is as follows: if f is any query function on n bits, then

$$M^{\mathrm{CC}}(f \circ \mathrm{IP}_{100 \log n}) = \Omega(M(f)),$$

where M is a query measure and M^{CC} is a corresponding communication measure. Sometimes, the stronger statement

$$M^{\mathrm{CC}}(f \circ \mathrm{IP}_{100 \log n}) = \Omega(M(f) \log n)$$

can be shown; note the we expect the $\log n$ factor to be there because we've added a $O(\log n)$-sized gadget to f, presumably increasing its complexity by a $\log n$ factor.

This type of lifting theorem cannot be used to give disjointness lower bounds in the model M^{CC}, and it can only give inner product lower bounds if the gadget is precisely the inner product gadget; hence it is much less useful than a lifting theorem with a constant-sized gadget. However, this type of lifting theorem can still be used to convert query separations to communication separations. For example, recently it has been shown there is a total query function f with $\mathrm{R}(f) = \Omega(\mathrm{Q}(f)^{2.66})$. Can we come up with a communication function F that has $\mathrm{R}^{\mathrm{CC}}(F) = \Omega(\mathrm{Q}^{\mathrm{CC}}(F)^{2.66})$? Well, if we had a lifting theorem for R with a $\log n$ size gadget, then we can lift f with that gadget G and get a lower bound on $\mathrm{R}^{\mathrm{CC}}(f \circ G)$; we could then argue that $\mathrm{Q}^{\mathrm{CC}}(f \circ G) = \tilde{O}(\mathrm{Q}(f))$, because Alice and Bob can compute f by mimicking the quantum query algorithm for f, and by computing the corresponding gadgets G whenever the query algorithm for f needs to read a bit of the input. This would let us conclude $\mathrm{R}^{\mathrm{CC}}(f \circ G) = \Omega(\mathrm{Q}^{\mathrm{CC}}(f \circ G)^{2.66})$ just by lifting the query separation, without having to do any extra work.

To get the above strategy to work, we actually only needed the gadget G to be computable using $O(\log n)$ communication, rather than needing it to be defined on $O(\log n)$ bits. There is therefore a third type of lifting theorem, which is even weaker: a lifting theorem with an index gadget that has $100 \log n$ bits of input to Alice and n^{100} bits on input to Bob. Lifting theorems with an index gadget can still be used to lift query separations to communication separations, but they are less satisfying.

As a final note, we mention that the index function with k bits on one side and 2^k bits on the other side contains all gadgets $G \colon \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ as sub-functions. This is because we can map from (x, y) to $(x, \psi(y))$ and vice versa, where $\psi(y)$ was defined in the proof of Lemma 9.1. Hence lifting with a $O(\log n)$ size gadget is strictly stronger than lifting with an index gadget.

Many measures, including both R^{CC} and D^{CC}, have been lifted with $O(\log n)$ sized gadgets, but not constant sized gadgets. Typically, the first proof uses an index gadget, and this gets improved to a $O(\log n)$ inner product gadget after a year or two. Other measures, primarily degree-like measures including deg, $\widetilde{\deg}$, and $\deg_{\mathrm{PP}}$, have been lifted with a constant sized gadget. Unfortunately, Q^{CC} has not been lifted with any gadget.

In the rest of this note, we will examine the communication analogues of the degree measures.

9.2 Approximate logrank

We've already seen the measure logrank, which was just defined as the logarithm of the rank of the communication matrix. We claimed that this measure corresponds to (exact) polynomial degree in query complexity.

We will now introduce an approximate version of logrank. Fix a communication function $F: \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, and consider its communication matrix. This is a $\{0,1\}$-valued matrix. We say that a real matrix A, with rows and columns indexed by $\mathcal{X}$ and $\mathcal{Y}$ respectively, *approximates* F to error ϵ if the following conditions hold:

1. $A[x,y] \in [0,1]$ for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$,
2. $|A[x,y] - F[x,y]| \leq \epsilon$ for all $(x,y) \in \mathrm{Dom}(F)$.

Note that this definition also generalizes to partial functions F; if $F[x,y] = *$, we simply require that the approximating matrix A satisfies $A[x,y] \in [0,1]$.

Actually, we will usually switch from the $\{0,1\}$ representation of a Boolean function to the $\{+1,-1\}$ representation. In the latter representation, the matrix F has entries in $\{+1,-1,*\}$, and any matrix A approximating F must satisfy $|A[x,y]| \leq 1$ for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$ and $|A[x,y] - F[x,y]| \leq 2\epsilon$ for all $(x,y) \in \mathrm{Dom}(F)$. Note that the allowed error increased to 2ϵ in the $\{+1,-1\}$ representation, since the values of F and A got stretched by a factor of 2. This is also what we did for polynomial approximations of a query function.

The approximate rank of a communication matrix F to error ϵ is the minimum rank of a matrix A which approximates F to error ϵ. Note that this can actually be slightly different when F is represented as a $\{0,1\}$-valued matrix and when it is represented as a $\{+1,-1\}$-valued matrix; converting between the two requires the transformations $J - 2F$ in one direction and $(J-F)/2$ in the other direction, where J is the all-1s matrix. Since $\mathrm{rank}(J) = 1$, these transformations may change the rank of F by at most 1. By default, we will represent F as a $\{+1,-1\}$ matrix, as this will turn out to be more convenient.

The approximate logrank of F is simply the log of the approximate rank of F. We denote this measure by $\log \widetilde{\mathrm{rank}}_\epsilon(F)$. It turns out that the error ϵ can be amplified; if we have a matrix A of rank k which approximates F to error ϵ, and if we have another target level $\epsilon' < \epsilon$, then there is a constant c which depends only on ϵ and ϵ' such that there must exist a matrix A' of rank at most k^c which approximates F to error ϵ'. Since $\log k^c = c \log k$, this is only a constant factor loss on the logarithm of the rank. We will not generally care about constant factors. This means that for our purposes, approximate logrank to any constant error $\epsilon \in (0, 1/2)$ is equivalent to any other constant error level. We will pick $\epsilon = 1/3$ to be the standard, and when $\epsilon = 1/3$, we omit it and simply write $\log \widetilde{\mathrm{rank}}(F)$.

Theorem 9.2. *For any (possibly partial) communication function F and any $\epsilon \in (0,1/2)$, we have $\mathrm{R}_\epsilon^{\mathrm{CC,PRIV}}(F) \geq \log(\widetilde{\mathrm{rank}}_\epsilon(F) - 1)$.*

Proof. Consider any randomized communication protocol which solves F to error ϵ. Let A be the matrix for which $A[x,y]$ is the probability that Alice and Bob output 1 in this protocol when given input (x,y). Since the protocol solves F to error ϵ against all inputs, we know that A approximates the $\{0,1\}$ version of F to error ϵ. We now claim that the rank of A is at most 2^T, where T is the number of rounds in this communication protocol.

Note that $A[x,y]$, the probability that Alice and Bob output 1, can be written as the sum over all transcripts $\Pi \in \{0,1\}^T$ of the probability that Alice and Bob output 1 and have exactly this transcript; here the *transcript* is the sequence of bits Alice and Bob sent to each other, in order. Let's denote by M_Π the matrix where $M_\Pi[x,y]$ is the probability that Alice and Bob communicate the transcript Π to each other, and then output 1. Then we have $A = \sum_{\Pi \in \{0,1\}^T} M_\Pi$. We now claim that each M_Π is a rank-1 matrix. To see this, note that since this is a private randomness

protocol, for any fixed (x, y) the probability that Alice sends the bits specified by Π is independent of the probability Bob sends the bits specified by Π. In other words, $M_\Pi[x, y]$ is equal to

$$\Pr[\text{Alice and Bob send bits specified by } \Pi] = \Pr[\text{Alice sends the bits specified by } \Pi] \cdot \Pr[\text{Bob sends the bits specified by } \Pi].$$

This means that $M_\Pi[x, y] = V_\Pi(x)U_\Pi(y)$ for some functions $V_\Pi(x)$ and $U_\Pi(y)$. Interpreting V_Π and U_Π as vectors, we can write $M_\Pi = V_\Pi U_\Pi^\dagger$. This means M_Π is rank-1, and hence A is a sum of 2^T rank-1 matrices, meaning that it has rank at most 2^T; this means the approximate rank of the $\{0, 1\}$ formulation of F is at most 2^T, so the approximate rank of the $\{+1, -1\}$ formulation is at most $2^T + 1$, as desired. □

It turns out that approximate logrank also lower bounds quantum communication complexity (without shared entanglement).

Theorem 9.3. *For any (possibly partial) communication function F and any $\epsilon \in (0, 1/2)$, we have $\mathrm{Q}_\epsilon^{\mathrm{CC}}(F) \geq \frac{1}{2}\log(\widetilde{\mathrm{rank}}_\epsilon(F) - 1)$, where Q^{CC} denotes the quantum communication complexity without shared entanglement.*

We will not go over the proof (see [BW01]). Note that this only applies to non-shared-entanglement quantum communication complexity; to lower bound the stronger shared entanglement model, we will need Theorem 9.6 in the next section.

One interesting corollary of these results is that $\widetilde{\mathrm{rank}}(I_n) = O(\mathrm{polylog}(n))$, even though we know $\mathrm{rank}(I_n) = n$ (here I_n denotes the $n \times n$ identity matrix). This is because we know that

$$\log \widetilde{\mathrm{rank}}(I_{2^k}) = \log \widetilde{\mathrm{rank}}(\mathrm{EQ}_k) \leq \mathrm{R}^{\mathrm{CC,PRIV}}(\mathrm{EQ}_k) = O(\mathrm{R}^{\mathrm{CC}}(\mathrm{EQ}_k) + \log k) = O(\log k),$$

and exponentienting both sides gives $\widetilde{\mathrm{rank}}(I_{2^k}) = O(\mathrm{poly}\, k)$, where $k = \log n$. This result may seem surprising, because we know that the identity matrix has rank n and it's hard to imagine an approximation matrix which is close to the identity matrix entry-wise but has rank logarithmic in n. Nevertheless, we know that such a matrix must exist.

9.3 Discrepancy and sign rank

Next, we define communication analogues of $\deg_{\mathrm{PP}}$ and $\deg_\pm$. The analogue of $\deg_\pm$ is called *sign rank*.

Definition 9.4. *The sign rank of a communication function, denoted $\mathrm{rank}_\pm(F)$, is the minimum rank of a real matrix A such that $A[x, y]F[x, y] > 0$ for all $(x, y) \in \mathrm{Dom}(F)$, where we use the $\{+1, -1\}$ version of F.*

Similar to how we showed that $\deg_\pm$ is equal to $\mathrm{R}_\pm$ and $\mathrm{Q}_\pm$ up to constant factors, ot is possible to show that $\log \mathrm{rank}_\pm(F)$ is equal to unbounded-error communication complexity (in the randomized or quantum models) up to constant factors, though we need additive constants as well as multiplicative constants (so things might break down a little bit when the sign rank is constant). Note that the definition of sign rank is equivalent to approximate rank where we take $\epsilon \to 1/2$. This measure is UPP in communication complexity.

We also define PP in communication complexity, which we call *discrepancy*. Instead of defining it in terms of the rank of an approximating matrix, though, we will define it in terms of an analogue of the dual of $\deg_{\mathrm{PP}}$ in query complexity.

Definition 9.5. *Let F be a (possibly partial) communication function. For a probability distribution μ over $\mathrm{Dom}(F)$, we define*

$$\mathrm{Disc}_\mu(F) \coloneqq \max_{u\in\{0,1\}^{|\mathcal{X}|},v\in\{0,1\}^{|\mathcal{Y}|}} |u^T(\mu \circ F)v|,$$

where we use F to denote the communication matrix in $\{+1,-1\}$ representation. We further define

$$\mathrm{Disc}(F) \coloneqq \min_\mu \mathrm{Disc}_\mu(F).$$

Note that $\mathrm{Disc}_\mu(F)$ is the maximum, over rectangles vu^T, of the difference between the 0-probability mass of μ in the rectangle and the 1-probability mass of μ in the rectangle. This can sort of be viewed as the correlation between F and the rectangle (under distribution μ), which is analogous to how in query complexity we wanted a distribution under which f has small correlation with every monomial.

Note that the discrepancy, as we've defined it, is smaller than 1; to get an actual communication measure, we generally look at

$$\log \frac{1}{\mathrm{Disc}(F)}.$$

We will soon try to relate some of these communication measures to each other and to randomized and quantum communication complexity, but first, we'll go on a long detour about a certain matrix norm called the gamma 2 norm.

9.4 The gamma 2 norm

An important matrix norm for the study of quantum computation is the gamma 2 norm. We've already briefly encountered a variant of it (called filtered gamma 2 norm) in the context of the negative-weight quantum adversary.

The gamma 2 norm satisfies the following important relationship with communication complexity.

Theorem 9.6. *Let Π be any T-round quantum communication protocol with shared entanglement which specifies Alice and Bob's communication and output when given input from $\mathcal{X}\times\mathcal{Y}$. Let $M \in [0,1]^{|\mathcal{X}|\times|\mathcal{Y}|}$ be the matrix where $M[x,y]$ is the probability that Alice and Bob output 1 on input (x,y). Then $\gamma_2(M) \le 2^T$.*

This theorem was shown by [LS09b], and we will not reproduce the proof here. Instead, we'll focus our attention on the study of γ_2 and its properties. It will turn out to be related to both discrepancy and approximate rank.

The gamma 2 norm can be viewed as a way of defining a "smoother" version of rank. The rank of a matrix, of course, is always an integer. The gamma 2 norm has many properties that make it similar to rank, but it is a *norm*, which means that if we scale a matrix by a factor of α, the gamma 2 norm would change by $|\alpha|$ (the rank would not change at all). In what follows, we roughly follow the presentation of [LSŠ08].

9.4.1 Definitions of the gamma 2 norm

We define the gamma 2 norm as follows. We call this definition the "min" formulation of the gamma 2 norm.

Definition 9.7. *Let A be an $m \times n$ real matrix. Then*

$$\gamma_2(A) \coloneqq \min_{B,C:BC=A} \|B\|_r \|C\|_c,$$

where the minimum is taken over pairs of matrices B and C such that $BC = A$, and where $\|B\|_r$ is the maximum 2-norm of a row of B and $\|C\|_c$ denotes the maximum 2-norm of a column of C.

This norm looks pretty unwieldy at first sight. It is not even clear why the minimum over B and C should be attained; is it a minimum, or an infimum? We now show that the gamma 2 norm can be written as the optimal solution of a semidefinite program. To do so, we will need the following notation. For an $m \times n$ matrix A, let $\hat{A}$ denote the $(m+n) \times (m+n)$ block matrix

$$\hat{A} \coloneqq \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}.$$

Let $J_{m,n}$ denote the $m \times n$ all-1s matrix. Note that the statement $BC = A$ is equivalent to the statement

$$(B^T \ \ C)^T (B^T \ \ C) = \begin{bmatrix} BB^T & A \\ A^T & C^T C \end{bmatrix},$$

and hence $BC = A$ if and only if $(B^T \ \ C)^T (B^T \ \ C) \circ \hat{J}_{m,n} = \hat{A}$, where $\circ$ denotes the Hadamard (entry-wise) product. Also note that when we minimize $\|B\|_r \|C\|_c$ in the gamma 2 norm of A, we can always assume without loss of generality that $\|B\|_r = \|C\|_c$, because otherwise we could multiply B by a constant and divide C by the same constant. Therefore, $\gamma_2(A)$ is equal to the minimum, over matrices $D = (B^T \ \ C)$ such that $D^T D \circ \hat{J}_{m,n} = \hat{A}$, of the maximum squared 2-norm of a column of D.

Since a matrix is positive semidefinite if and only if it can be written as $D^T D$ for some matrix D, we can write $\gamma_2(A)$ as the minimum, over positive semidefinite $(m+n) \times (m+n)$ matrices X such that $X \circ \hat{J}_{m,n} = \hat{A}$, of the maximum diagonal entry of X. That is,

$$\begin{aligned} \gamma_2(A) = \quad & \min & \eta & & \\ & \text{s.t.} & X_{ii} & \leq \eta & \forall i \in [m+n] \\ & & X \circ \hat{J}_{m,n} & = \hat{A} & \\ & & X & \succeq 0. & \end{aligned}$$

Now that we have written $\gamma_2(A)$ as the optimal solution to a semidefinite program, we can take the dual. We also need to check that strong duality holds (as it usually does), but we skip over that for this presentation. Then we have

$$\begin{aligned} \gamma_2(A) = \quad & \max & \langle Q, \hat{A} \rangle & \\ & \text{s.t.} & \Delta \circ I & = \Delta \\ & & \langle \Delta, I \rangle & = 1 \\ & & \Delta & \succeq Q \\ & & Q \circ \hat{J}_{m,n} & = Q. \end{aligned}$$

Note that Δ is diagonal, and that all matrices are $(m+n) \times (m+n)$. Since $\Delta \succeq Q$, we have $\Delta \geq 0$. Now, note that $\Delta \succeq Q$ is equivalent to $u^T(\Delta - Q)u \geq 0$ for all u. For any positive vector $\beta > 0$, this is equivalent to $(\beta \circ u)^T(\Delta - Q)(\beta \circ u) \geq 0$, since quantifying over all vectors u is the same as quantifying over all vectors $\beta \circ u$ when β has nonzero entries. The latter is equivalent to $u^T((\Delta - Q) \circ \beta\beta^T)u \geq 0$, which is equivalent to $\Delta \circ \beta\beta^T \succeq Q \circ \beta\beta^T$.

We will restrict to $\Delta[i,i] > 0$ for all i for now, and take the supremum instead of the maximum; since in the limit, we get take diagonal entries of Δ to 0, this supremum will equal the previous objective value. We then pick $\beta[i] = 1/\sqrt{\Delta[i,i]}$. Setting $Q' = \beta\beta^T \circ Q$, the condition $\Delta \succeq Q$ becomes equivalent to the condition $I \succeq Q'$, or equivalently $\|Q'\| \leq 1$, where $\|\cdot\|$ is the spectral (operator) norm. The constraint $Q \circ \hat{J}_{m,n} = Q$ becomes equivalent to $Q' \circ \hat{J}_{m,n} = Q'$. Let v be the vector $v[i] = 1/\beta[i]$. Then the objective value becomes $\langle Q' \circ vv^T, \hat{A}\rangle = v^T(Q' \circ \hat{A})v$, and the constraint $\langle \Delta, I\rangle = 1$ becomes $\langle vv^T, I\rangle = 1$ or $\|v\|_2 = 1$. The maximum of $v^T(Q' \circ \hat{A})v$ over unit vectors v is the definition of $\|Q' \circ \hat{A}\|$, so we can replace the objective value with the latter expression and remove the constraint $\|v\|_2 = 1$ and the variable v.

We've concluded that $\gamma_2(A)$ is the maximum, over symmetric matrices matrices Q' satisfying $Q' \circ \hat{J}_{m,n} = Q'$ and $\|Q'\| \leq 1$, of $\|Q' \circ \hat{A}\|$. Since Q' is symmetric and $Q' \circ \hat{J}_{m,n} = Q'$, we can write

$$Q' = \begin{bmatrix} 0 & P \\ P^T & 0 \end{bmatrix}$$

for some matrix P, and then $\|Q' \circ \hat{A}\| = \|P \circ A\|$ and $\|Q'\| = \|P\|$. Hence $\gamma_2(A)$ is equal to the maximum, over $m \times n$ matrices P satisfying $\|P\| \leq 1$, of $\|P \circ A\|$. By scaling, we can conclude

$$\gamma_2(A) = \max_{P \neq 0} \frac{\|P \circ A\|}{\|P\|},$$

where $\|\cdot\|$ denotes the spectral norm. We call this the "max" formulation of the gamma 2 norm.

9.4.2 Basic properties of the gamma 2 norm

We collect a few other useful properties of the gamma 2 norm, which we compare with rank.

Lemma 9.8. *Let A and B be real $m \times n$ matrices, and let $\lambda \in \mathbb{R} \setminus \{0\}$. Then the following hold:*

1. $\gamma_2(A+B) \leq \gamma_2(A) + \gamma_2(B)$, $\operatorname{rank}(A+B) \leq \operatorname{rank}(A) + \operatorname{rank}(B)$.
2. $\gamma_2(\lambda A) = |\lambda|\gamma_2(A)$, $\operatorname{rank}(\lambda A) = \operatorname{rank}(A)$.
3. $\gamma_2(A) = 0 \Leftrightarrow A = 0 \Leftrightarrow \operatorname{rank}(A) = 0$.
4. $\gamma_2(J) = \operatorname{rank}(J) = 1$ *if J is an all-1 matrix of any dimension.*
5. $\gamma_2(A^T) = \gamma_2(A)$ *and* $\operatorname{rank}(A^T) = \operatorname{rank}(A)$.
6. $\gamma_2(A)$ *is invariant under rearranging, duplicating, or negating rows or columns of A.* $\operatorname{rank}(A)$ *is invariant under these, and also under scaling rows or columns of A by a non-zero scalar.*
7. *If C is a submatrix of A, then* $\gamma_2(C) \leq \gamma_2(A)$ *and* $\operatorname{rank}(C) \leq \operatorname{rank}(A)$.
8. $\gamma_2(A \otimes B) = \gamma_2(A)\gamma_2(B)$ *and* $\operatorname{rank}(A \otimes B) = \operatorname{rank}(A)\operatorname{rank}(B)$, *where $\otimes$ is the Kronecker (tensor) product.*
9. $\gamma_2(A \circ B) \leq \gamma_2(A)\gamma_2(B)$ *and* $\operatorname{rank}(A \circ B) \leq \operatorname{rank}(A)\operatorname{rank}(B)$, *where $\circ$ is the Hadamard (entry-wise) product.*
10. $\|A\|_\infty \leq \gamma_2(A) \leq \|A\|_\infty \sqrt{\operatorname{rank}(A)}$, *where $\|A\|_\infty$ is the largest absolute value of an entry in A.*
11. $\gamma_2(A)$ *is a continuous function of A, whereas* $\operatorname{rank}(A)$ *is an integer-valued function of A.*

The first three properties say that γ_2 is a norm. It shares many properties with rank, though of course, rank is invariant under scaling by λ whereas $\gamma_2(\lambda A) = |\lambda|\gamma_2(A)$. For this reason, it is sometimes better to think of $\gamma_2(A)$ as similar to (but smaller than) $\|A\|_\infty \cdot \text{rank}(A)$, that is, the rank of A multiplied by the largest entry of A.

Proof. (1) follows because there is some matrix P such that $\|P\| = 1$ and $\gamma_2(A+B) = \|(A+B)\circ P\|$, and then $\gamma_2(A+B) = \|A\circ P + B\circ P\| \le \|A\circ P\| + \|B\circ P\| \le \gamma_2(A) + \gamma_2(B)$. (2), (3), and (4) follow easily from $\gamma_2(A) = \max_{P\neq 0} \|A\circ P\|/\|P\|$. (5) follows from the fact that the spectral norm stays the same when we take the transpose of a matrix. (6) follows from the fact that rearranging and negating rows or columns does not change the spectral norm $\|\cdot\|$; it remains to handle duplicating rows or columns. Due to (5), it suffices to handle duplicating rows. Since rearranging doesn't change $\gamma_2(A)$, we can assume the first row is duplicated. It is clear that this cannot decrease $\gamma_2(A)$, because we could always take P to have an all-0 row on the duplicate row. This also cannot increase $\gamma_2(A)$ due to its characterization as $\min_{BC=A} \|B\|_r\|C\|_c$, because we could always duplicate the corresponding row of B. (7) follows by setting the rows and columns of P that are outside of C to be all 0.

For (8), the lower bound $\gamma_2(A\otimes B) \ge \gamma_2(A)\gamma_2(B)$ follows by using the max characterization of γ_2 and taking P for $A\otimes B$ to be $P_A \otimes P_B$, where P_A and P_B maximize $\|A\circ P_A\|/\|P_A\|$ and $\|B\circ P_B\|/\|P_B\|$ respectively. The upper bound $\gamma_2(A\otimes B) \le \gamma_2(A)\gamma_2(B)$ follows by using the min characterization of γ_2, writing $A = CD$ and $B = EF$ such that $\|C\|_r\|D\|_c = \gamma_2(A)$ and $\|E\|_r\|F\|_c = \gamma_2(B)$, and then taking the matrices $C\otimes E$ and $D\otimes F$. The Kronecker product satisfies $(C\otimes E)(D\otimes F) = (CD)\otimes(EF) = A\otimes B$, so these matrices are feasible for the minimization version of $\gamma_2(A\otimes B)$. The 2-norm of each row of $C\otimes E$ is the product of a 2-norm of a row of C and a 2-norm of a row of E, and the 2-norm of each column of $D\otimes F$ is the product of a 2-norm of a column of D and a 2-norm of a column of F. From this it follows that $\gamma_2(A\otimes B) \le \gamma_2(A)\gamma_2(B)$.

(9) follows from observing that the Hadamard product $A\circ B$ is a submatrix of the Kronecker product $A\otimes B$. For (10), the inequality $\gamma_2(A) \ge \|A\|_\infty$ follows from the max characterization of $\gamma_2(A)$, picking P to be all-0s except for a single 1 entry at the coordinate where the largest entry of A is. We prove the other inequality in Section 9.4.4 (based on [LSŠ08]). Finally, (11) follows from the fact that $\gamma_2(A+\epsilon B) \le \gamma_2(A) + \epsilon\gamma_2(B) \le \gamma_2(A) + \epsilon\|B\|_\infty\sqrt{\text{rank}(B)}$, so as $\epsilon \to 0$, we have $\gamma_2(A+\epsilon B) \to \gamma_2(A)$. □

9.4.3 A quick detour about other matrix norms

We review some other matrix norms. There are many, many of these norms, which can get quite confusing; we'll categorize them into three primary categories: norms which treat matrices as vectors, operator norms, and Schatten norms. The γ_2 norm is in none of these categories.

Vector p-norms. One way of getting a norm is to treat the matrix as a vector, and take some p-norm (recall that the p-norm of a vector v, denoted $\|v\|_p$, is defined as $(\sum_i |v_i|^p)^{1/p}$). This would depend on the absolute values of the entries of the matrix, but not on their arrangement inside the matrix or their signs. Common values of p are $p = 1, 2, \infty$; recall that $p = \infty$ means the largest absolute value of an entry. Any value of $p \in [1,\infty]$ gives a norm.

Duality of norms. Before we proceed to discuss other matrix norms, we will define the *dual* of the norm $\|\cdot\|_*$ by

$$\|v\|_{dual} := \max_{u:\|u\|_*\le 1} |\langle u, v\rangle| = \max_{u\neq 0} \frac{|\langle u, v\rangle|}{\|u\|_*}.$$

This is the maximum inner product v can have with a vector of norm 1, where the "norm 1" is measured according to the original vector norm. Note that by interpreting the inner product $\langle A, B\rangle$ of matrices in the natural way, this lets us take the dual of any matrix norm. The dual norm of the dual norm always gives back the original norm. It turns out that the dual of the vector p-norm is a vector p'-norm, where $1/p' + 1/p = 1$; that is, $p' = p/(p-1)$. This is called the conjugate of p. In particular, the dual of the 2-norm is the 2-norm, the dual of the 1-norm is the ∞-norm, and the dual of the ∞-norm is the 1-norm.

Operator norms. Another way to define a norm of a matrix is to treat the matrix as an *operator*, that is, a function from vectors to vectors. Then we can pick a norm on the input vectors and a different norm on the output vectors, and define the norm of the matrix as the largest amount by which it might "stretch" an input vector. That is, we define

$$\|A\|_{p\to q} \coloneqq \sup_{v\neq 0} \frac{\|Av\|_q}{\|v\|_p}.$$

Equivalently, by scaling v so that $\|v\|_p = 1$, this is the maximum norm of a vector A may output when given a vector of norm 1 as input. Note that we allow the input space to be measured by a possibly different norm than the output space; we measure the input vector v using a p-norm, and we measure the output vector Av using a q-norm. The resulting measure of A turns out to be a norm, which we call the "p to q operator norm". Since $\|Av\|_q = \max_{u:\|u\|_{q/(q-1)}\leq 1}\langle u, Av\rangle$, the $p \to q$ norm can be written

$$\|A\|_{p\to q} = \max_{u,v:\|u\|_{q'}\leq 1,\|v\|_p\leq 1} u^T Av,$$

where $q' = q/(q-1)$ is the conjugate of q.

Note that the spectral norm, which we've previously denoted by $\|A\|$, is the 2-to-2 operator norm $\|A\|_{2\to 2}$. For any p, the $1 \to p$ norm is the maximum p-norm of a column, because maximizing over vectors with $\|v\|_1 = 1$ turns out to be the same as maximizing over basis vectors e_i. Similarly, the $p \to \infty$ norm is the maximum p'-norm of a row, where $p' = p/(p-1)$ is the conjugate of p. In particular, the $1 \to \infty$ norm is just $\|A\|_\infty$, the maximum absolute value of an entry of A. The dual norm of an operator norm does not have a closed form in general; it does not have to be an operator norm. Another property is that $\|A^T\|_{p\to q} = \|A\|_{q'\to p'}$, where q' and p' are the conjugates of q and p.

Let's go over all values of $p, q \in \{1, 2, \infty\}$. When $p = 1$ or $q = \infty$, the $p \to q$ norm is just the maximum norm of a row or column of A, as previously described. When they are both 2, we get the spectral norm. The remaining cases are $\infty \to 1$, $\infty \to 2$, and $2 \to 1$. Note that $\|A\|_{\infty\to 2} = \|A^T\|_{2\to 1}$, so we can convert from one to the other by taking the transpose; we've essentially reduced to just two norms we don't understand, the $\infty \to 1$ norm and the $2 \to 1$ norm. They don't have a closed-form representation (they are NP-hard to compute), but we will discuss the $\infty \to 1$ norm in detail later, as it turns out to be closely related to the γ_2 norm. Also note that $p \to q$ norms are increasing in p and decreasing in q, so the largest operator norm is $\infty \to 1$ and the smallest is $1 \to \infty$ (which is $\|A\|_\infty$).

Schatten norms. There is yet a third popular way of defining the norm of a matrix: this is to take the singular values $\sigma_1, \sigma_2, \ldots, \sigma_k$ of the matrix, and apply some p-norm to the vector $(\sigma_1, \sigma_2, \ldots, \sigma_k)$. This third way is called the Schatten p-norm of a matrix. The notation $\|A\|_p$ is therefore ambiguous, because it might denote either the Schatten p-norm, or the p-norm of A treated as a vector, or sometimes the $p \to p$ operator norm. The important cases of the Schatten

p-norms are $p = 1, 2, \infty$. The $p = 1$ case is called the *trace norm*. The $p = 2$ case is called the *Frobenius norm*, and it turns out to equal $\|A\|_2$, the 2-norm of A treated as a vector. The case $p = \infty$ gives the spectral norm $\|A\| = \|A\|_{2\to 2}$. Hence out of $p \in \{1, 2, \infty\}$, only $p = 1$ gave a new norm we didn't see before, which is the fairly important trace norm.

The trace norm, which is the sum of singular values, is the dual norm of the spectral norm. This is because the dual of a Schatten p-norm is the Schatten p'-norm, with $1/p' + 1/p = 1$. The Frobenius norm is its own dual.

We note that the Schatten norms can give us a lower bound on the rank. The reason is that using Cauchy-Schwartz on the trace norm $\sigma_1 + \sigma_2 + \cdots + \sigma_k$, we get that it is at most $\sqrt{k(\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_k^2)}$, which means that the trace norm of a matrix is at most $\sqrt{k}$ times the Frobenius norm. The value k is the number of non-zero singular values, which is the rank; hence $\text{rank}(A) \geq \|A\|_{tr}^2/\|A\|_F^2$, where $\|\cdot\|_{tr}$ is the trace norm and $\|\cdot\|_F$ is the Frobenius norm.

9.4.4 More properties of the gamma 2 norm

The γ_2 norm is not any of the norms mentioned in the previous section. In its min formulation, it is

$$\gamma_2(A) = \min_{B,C:BC=A} \|B\|_{2\to\infty}\|C\|_{1\to 2}.$$

In its max formulation, it is $\gamma_2(A) = \max_{Q:\|Q\|=1} \|Q \circ A\|$. We can apply the dual definition of the spectral norm: since it is the dual of the trace norm, it satisfies

$$\|Q\| = \max_{B\neq 0} \frac{\langle B, Q\rangle}{\|B\|_{tr}}$$

$$\|Q \circ A\| = \max_{C\neq 0} \frac{\langle C, Q \circ A\rangle}{\|C\|_{tr}},$$

and hence

$$\gamma_2(A) = \max_{Q,C\neq 0} \min_{B\neq 0} \frac{\langle Q, C \circ A\rangle \|B\|_{tr}}{\|C\|_{tr}\langle B, Q\rangle}.$$

Note that we could always take $B = C \circ A$, in which case we would get $\|C \circ A\|_{tr}/\|C\|_{tr}$; hence $\gamma_2(A) \leq \max_{C\neq 0} \|C \circ A\|_{tr}/\|C\|_{tr}$. Using the dual characterization of $\|C \circ A\|_{tr}$ and of $\|C\|_{tr}$, we can similarly show that $\max_{C\neq 0} \|C \circ A\|_{tr}/\|C\|_{tr} \leq \max_{Q\neq 0} \|Q \circ A\|/\|Q\|$. Hence we have a new characterization for the gamma 2 norm:

$$\gamma_2(A) = \max_{Q\neq 0} \frac{\|Q \circ A\|_{tr}}{\|Q\|_{tr}}.$$

Now, consider the matrix Q which maximizes this expression, and write its singular value decomposition $Q = \sigma_1 v_1 u_1^T + \sigma_2 v_2 u_2^T + \cdots + \sigma_k v_k u_k^T$, where the v_i vectors are orthonormal and the u_i vectors are orthonormal. Then $\|Q\|_{tr} = \sigma_1 + \cdots + \sigma_k = \sigma_1\|v_1 u_1^T\|_{tr} + \cdots + \sigma_k\|v_k u_k^T\|_{tr}$. Moreover, since the trace norm is a norm, we have $\|Q \circ A\|_{tr} \leq \sigma_1\|v_1 u_1^T \circ A\|_{tr} + \cdots + \sigma_k\|v_k u_k^T \circ A\|_{tr}$. Since a weighted average of a list is less than or equal to the maximum, we have

$$\frac{\|Q \circ A\|_{tr}}{\|Q\|_{tr}} \leq \max_i \frac{\|v_i u_i^T \circ A\|_{tr}}{\|v_i u_i^T\|_{tr}},$$

and hence the maximum over Q is always attained for a rank-1 matrix vu^T. Hence we actually have

$$\gamma_2(A) = \max_{u,v:\|u\|_2=\|v\|_2=1} \|vu^T \circ A\|_{tr}.$$

We call this the trace norm formulation of the gamma 2 norm.

We saw the inequality $\|B\|_{tr} \leq \|B\|_F \sqrt{\operatorname{rank} B}$, and hence we have $\gamma_2(A) \leq \max_{u,v:\|u\|=\|v\|=1} \|A \circ vu^T\|_F \sqrt{\operatorname{rank}(A \circ vu^T)}$. Note that $\|A \circ vu^T\|_F^2 = \sum_{ij} A_{ij}^2 v_i^2 u_j^2 \leq \sum_{ij} \|A\|_\infty^2 v_i^2 u_j^2 = \|A\|_\infty^2$, and hence $\|A \circ vu^T\|_F \leq \|A\|_\infty$. Also, it is not hard to see that $\operatorname{rank}(A \circ vu^T) \leq \operatorname{rank}(A)$, and hence we get $\gamma_2(A) \leq \|A\|_\infty \sqrt{\operatorname{rank}(A)}$, completing the proof of Lemma 9.8.

9.4.5 The dual norm of the gamma 2 norm

It will also be important to study the dual norm of the gamma 2 norm, which we denote $\gamma_2^*(A)$. By definition of a dual norm, we have

$$\gamma_2^*(A) = \max_{B:\gamma_2(B)=1} \langle B, A \rangle,$$

and since the dual norm of a dual norm is the original norm, we have

$$\gamma_2(A) = \max_{B:\gamma_2^*(B)=1} \langle B, A \rangle.$$

Recall that we've already seen such a definition of $\gamma_2(A)$: the dual of the semidefinite program was

$$\begin{aligned} \gamma_2(A) = \quad \max \quad & \langle Q, \hat{A} \rangle & \\ \text{s.t.} \quad & \Delta \circ I & = \Delta \\ & \langle \Delta, I \rangle & = 1 \\ & \Delta & \succeq Q \\ & Q \circ \hat{J}_{m,n} & = Q. \end{aligned}$$

In the above program, Q is a symmetric matrix satisfying $Q \circ \hat{J}_{m,n}$, which means that $Q = \hat{B}$ for some matrix B. Hence $\langle Q, \hat{A} \rangle = 2\langle B, A \rangle$. Letting $C = 2B$, we see that $\gamma_2(A)$ is the maximum over matrices C of $\langle C, A \rangle$, subject to $\hat{B} \preceq \Delta$ where Δ is a diagonal matrix which sums to 2. Since $\gamma_2(A)$ is also the maximum over matrices C of $\langle C, A \rangle$ subject to $\gamma_2^*(C) = 1$, we conclude that $\gamma_2^*(C)$ is the minimum value of λ such that $\hat{C} \preceq \Delta$, where Δ is a diagonal matrix with $\sum_i \Delta[i,i] \leq 2\lambda$. That is,

$$\gamma_2^*(A) = \min \left\{ \frac{\langle I, \Delta \rangle}{2} : \Delta \succeq \hat{A}, \Delta \text{ diagonal} \right\}.$$

We can also get a formulation of $\gamma_2^*(A)$ in terms of the Frobenius norm of a factorization of A (recall that the Frobenius norm is the vector 2-norm of the matrix).

Lemma 9.9. *For any real matrix A,*

$$\gamma_2^*(A) = \min \|B\|_F \|C\|_F,$$

where the minimum is taken over matrices B and C such that $BC = A$, B has orthogonal rows, and C has orthogonal columns.

Proof. Let Δ be a diagonal matrix with $\langle I, \Delta \rangle / 2 = \gamma_2^*(A)$ and $\Delta \succeq \hat{A}$. Then we have $\Delta - \hat{A} \succeq 0$, so we can write $\Delta - \hat{A} = M^T M$ for some matrix M. If A is an $m \times n$ matrix, we can write $M = [-B^T \quad C]$, where B^T constitutes the first m columns of M and C constitutes the last n columns of M; then since $M^T M = \Delta - \hat{A}$, we have $BC = A$, and BB^T and $C^T C$ must be diagonal matrices which are diagonal blocks of Δ (in particular, B and C must have orthogonal rows and columns, respectively). Hence the squared singular values of B are just the diagonal entries of

B^TB, and similarly for C; this means that $\|B\|_F^2 + \|C\|_F^2 = \langle I, \Delta\rangle$ (using the Schatten 2-norm interpretation of the Frobenius norm), so $(\|B\|_F^2 + \|C\|_F^2)/2 = \gamma_2^*(A)$. We can now scale B by a constant and divide C by the same constant in order to get $\|B\|_F = \|C\|_F$; this does not affect the identity $A = BC$ or the orthogonality of the rows and columns. Hence the minimum of $\|B\|_F\|C\|_F$ at most $\gamma_2^*(A)$.

In the other direction, suppose we have $A = BC$ where B has orthogonal rows and C has orthogonal columns. By scaling, we can assume $\|B\|_F = \|C\|_F$. Then let $M = [-B^T \quad C]$, and observe that M^TM is a matrix which looks like $\Delta - \hat{A}$ for a diagonal matrix Δ. Hence $\Delta - \hat{A} \succeq 0$, and $\langle \Delta, I\rangle/2 = (\|B\|_F^2 + \|C\|_F^2)/2 = \|B\|_F\|C\|_F$, from which the desired result follows. □

While we don't prove this here, the trace norm has a very similar formulation to the above.

Lemma 9.10. *For any real matrix A,*

$$\|A\|_{tr} = \min_{B,C:BC=A} \|B\|_F\|C\|_F.$$

The difference between the trace norm and γ_2^* is therefore just the requirement that the rows of B and columns of C are orthogonal; hence for all A, we have

$$\|A\|_{tr} \leq \gamma_2^*(A),$$

and taking duals,

$$\gamma_2(A) \leq \|A\|.$$

9.4.6 Grothendieck's inequality

An important and highly nontrivial property of the γ_2 norm is something known as Grothendieck's inequality.

Theorem 9.11 (Grothendieck)**.** *There is a universal constant K_G such that for any real matrix A,*

$$\|A\|_{\infty\to 1} \leq \gamma_2^*(A) \leq K_G\|A\|_{\infty\to 1}.$$

The constant K_G is known as Grothendieck's constant. The best value of K_G is not known, though it is known to be between 1.67 and 1.79.

Up to constant factors, this gives us a completely different interpretation of $\gamma_2^*(A)$: it is simply the $\infty \to 1$ operator norm, which is the largest of the operator norms. It's worth saying a few words about this operator norm. Recall that it can be defined

$$\|A\|_{\infty\to 1} = \max_{u,v:\|u\|_\infty = \|v\|_\infty \leq 1} u^TAv.$$

Note that the only constraints on u and v are the linear constraints $-1 \leq u_i, v_j \leq 1$ for all i and j; hence v and u are constrained to be in a box. It's not hard to see that the optimal choices of u and v will be at some corner of the box, so we can take $u \in \{+1, -1\}^m$, $v \in \{+1, -1\}^n$, and get

$$\|A\|_{\infty\to 1} = \max_{u\in\{+1,-1\}^m, v\in\{+1,-1\}^n} \langle A, vu^T\rangle.$$

Equivalently, $\|A\|_{\infty\to 1}$ is the largest value we can get by negating some set of rows S and some set of columns T of A, and then summing up the entries of the resulting matrix. In particular, this will always be at most $\|A\|_1$ (the sum of absolute values of the entries of A), and if $A \geq 0$, then $\|A\|_{\infty\to 1} = \|A\|_1$.

The dual norm of the $\infty \to 1$ operator norm is something known as the *nuclear norm*, which we denote $\nu(A)$. If A is $m \times n$, it can be written

$$\nu(A) = \min \left\{ \sum_{i=1}^{k} \Delta_{ii} : X\Delta Y = A, X \in \{+1,-1\}^{m\times k}, Y \in \{+1,-1\}^{k\times n}, \Delta \geq 0, \Delta \text{ diagonal} \right\}.$$

Since $\nu(A)$ is the dual norm of $\|A\|_{\infty\to 1}$ (something we don't prove here), we can take the dual norms of the norms in Grothendieck's inequality. Note that if some pair of norms satisfies $\|A\|_{n1} \leq \|A\|_{n2}$ for all A, then their dual norms $n1^*$ and $n2^*$ satisfy $\|A\|_{n1^*} \geq \|A\|_{n2^*}$ for all A. This is because

$$\|A\|_{n1^*} = \max_{B:\|B\|_{n1}\leq 1} \langle A, B\rangle \geq \max_{B:\|B\|_{n2}\leq 1} \langle A, B\rangle = \|A\|_{n2^*}.$$

Also note that while the dual norm of $\|\cdot\|_{\infty\to 1}$ is $\nu(\cdot)$, if we define $K_G \cdot \|\cdot\|_{\infty\to 1}$ as its own norm, its dual would be $\nu(\cdot)/K_G$. Hence Grothendieck's inequality is equivalent to the following.

Corollary 9.12. *For any real matrix A,*

$$\gamma_2(A) \leq \nu(A) \leq K_G \cdot \gamma_2(A),$$

where K_G is Grothendieck's constant, $1.67 \leq K_G \leq 1.79$.

Up to constant factors, this gives us yet another interpretation for the magical norm $\gamma_2(\cdot)$.

9.4.7 The cut norm and the mu norm

We define yet another matrix norm, called the cut norm, which is closely related to the $\infty \to 1$ operator norm.

Definition 9.13. *Let A be a real matrix. The cut norm of A, denoted $\|A\|_C$, is the maximum value of $|\langle R, A\rangle|$ over all rectangles R, where recall that a rectangle is $R = vu^T$ for $\{0,1\}$-valued vectors v and u.*

The cut norm is important for our purposes, as it comes up in the definition of discrepancy: by definition, we have $\text{Disc}_\mu(F) = \|\mu \circ F\|_C$, and hence $\text{Disc}(F) = \min_\mu \|\mu \circ F\|_C$, where the minimum is over matrices μ which are probability distributions (i.e. $\mu \geq 0$ and the entries of μ sum to 1).

The dual of the cut norm is sometimes called the mu norm.

Definition 9.14. *Let A be a real matrix. The mu norm of A, denoted $\mu(A)$, is the minimum value of $\sum_{i=1}^{k} |w_i|$ over real numbers $w_1, w_2, \ldots, w_k \in \mathbb{R}$ which satisfy $A = w_1R_1 + w_2R_2 + \ldots w_kR_k$ for rectangles $R_1, R_2, \ldots, R_k$.*

We claim that the cut norm and the mu norm are duals.

Lemma 9.15. *The dual of the cut norm is the mu norm.*

Proof. The dual of the cut norm, applied to a matrix A, is $\max_{B:\|B\|_C\leq 1}\langle B, A\rangle$. We can write this as a linear program, where the variables are the entries of B, the objective value is $\langle B, A\rangle$, and the constraints are $\langle B, R\rangle \leq 1$ and $-\langle B, R\rangle \leq 1$ for each rectangle R. The dual of this linear program is the minimum of $\sum_R y_R + z_R$ subject to $\sum_R y_R R - \sum_R z_R R = A$, and $y_R, z_R \geq 0$ for all R. It is clear that in the optimal solution, we won't have $y_R > 0$ and $z_R > 0$ for the same rectangle R. Hence setting $w_R = y_R - z_R$, the program becomes the minimum of $\sum_R |w_R|$ subject to $\sum_R w_R R = A$, which is the mu norm. □

We now show that the cut norm is equivalent to γ_2^*, and the mu norm is equivalent to γ_2.

Theorem 9.16. *For every real matrix A,*

$$\|A\|_C \leq \gamma_2^*(A) \leq 4K_G\|A\|_C,$$

$$\gamma_2(A) \leq \mu(A) \leq 4K_G\gamma_2(A),$$

where $1.67 \leq K_G \leq 1.79$ is Grothendieck's constant. In particular, $4K_G \leq 8$.

Proof. We only need to show the first line, relating γ_2^* to the cut norm $\|\cdot\|_C$; the rest will follow from duality. Additionally, we appeal to Grothendieck's inequality, and show only that $\|A\|_C \leq \|A\|_{\infty\to 1} \leq 4\|A\|_C$.

We start with the right hand side. Recall that $\|A\|_{\infty\to 1} = \max_{u,v:\|u\|_\infty=\|v\|_\infty=1} u^TAv$, and that we can always assume the maximum over u and v will occur for $\{+1,-1\}$-valued vectors (this is because u^TAv is either increasing or decreasing in each entry of u or v, when that entry is considered separately from the rest; hence we can increase or decrease that entry until it reaches 1 or -1 and only increase u^TAv, without affecting $\|u\|_\infty$ or $\|v\|_\infty$). Now, $u^TAv = \langle vu^T, A\rangle$, and vu^T is a matrix made up of two all-ones rectangles and two all-minus-ones rectangles; hence vu^T can be written $R_1 + R_2 - R_3 - R_4$ for rectangles R_1, R_2, R_3, R_4. Hence $|\langle vu^T, A\rangle| \leq 4\max_R |\langle R, A\rangle| = 4\|A\|_C$.

To see that $\|A\|_C \leq \|A\|_{\infty\to 1}$, consider the rectangle $R_1 = vu^T$ for which $\|A\|_C = |\langle A, R_1\rangle|$. Consider the three complementary rectangles R_2, R_3, R_4 that we can get by flipping the 0s and 1s in u, or in v, or in both. Take all the combinations of writing $R_1 + R_i - R_j - R_k$ where $\{1,i,j,k\} = \{1,2,3,4\}$. There are three such combinations, and gives a matrix which can be written ba^T for sign vectors a and b. Adding them together gives $3R_1 - R_2 - R_3 - R_4$. Adding J as well gives $4R_1$. Hence we've represented $4R_1$ as the sum of 4 matrices of the form ba^T for sign vectors a and b; this means that $4|\langle R_1 A\rangle| \leq 4\|A\|_{\infty\to 1}$, as desired. □

We note that this mu norm gives us the most intuitive version of the gamma 2 norm we've seen so far. Of course, the rank of a matrix A is the minimum number k such that A can be written as a sum of k rank-1 matrices. The gamma 2 norm of A is different from the rank in two ways: first, instead of representing A as a sum of rank-1 matrices, it represents A as a (weighted) sum of rectangles. However, rectangles aren't too different from rank-1 matrices in practice. Second, instead of counting *how many* rectangles were used, we measure the total weight assigned to the rectangles.

9.5 Approximate gamma 2 norm and logrank

We now define a new complexity measure called the approximate gamma 2 norm, which we'll see is closely related to the approximate rank.

Definition 9.17. *Let F be a (possibly partial) communication function. The approximate gamma 2 norm of F to error ϵ, denoted $\tilde{\gamma}_{2,\epsilon}(F)$, is the minimum value of $\gamma_2(A)$ over matrices A which ϵ-approximate F in the $\{+1,-1\}$ representation. In other words, it is the minimum of $\gamma_2(A)$ over matrices A with entries in $[-1,1]$ and satisfying $|A[x,y] - F[x,y]| \leq 2\epsilon$ whenever $F[x,y] \neq *$.*

When $\epsilon = 1/3$, we omit it and write $\tilde{\gamma}_2(F)$.

Note that the set of all matrices A which ϵ-approximate F is compact, and since $\gamma_2(\cdot)$ is continuous, we know that this minimum is attained; in other words, we our use of minimum in the above definition did not need to be an infimum, and for every F and $\epsilon \in [0,1/2)$ there is an ϵ-approximating matrix A such that $\gamma_2(A) = \tilde{\gamma}_{2,\epsilon}(F)$.

Now, recall that if a quantum communication protocol computes F to error ϵ, then the matrix M of acceptance probabilities associated with that protocol must ϵ-approximate F in the $\{0,1\}$ representation. This gives the following corrolary.

Corollary 9.18. *For any (possibly partial) communication function F, the quantum communication complexity with shared entanglement of F satisfies $\mathrm{Q}_\epsilon^{\mathrm{CC},*}(F) \geq \log \tilde{\gamma}_{2,\epsilon}(F) - 1$.*

Proof. Consider a such a protocol Π for F, and let A be the matrix of its acceptance probabilities. Then A ϵ-approximates F in the $\{0,1\}$ representation, and by Theorem 9.6, $\gamma_2(A) \leq 2^{\mathrm{Q}^{\mathrm{CC},*}(F)}$. Now consider the negated protocol on which Alice and Bob run Π but output the opposite value. The matrix of acceptance probabilities of this negated protocol is $J - A$, and since this protocol has the same quantum communication cost, we have $J - A \leq 2^{\mathrm{Q}^{\mathrm{CC},*}(F)}$. Note that $J - 2A$ is a matrix which ϵ-approximates F. Also,

$$\gamma_2(J - 2A) \leq \gamma_2(J - A) + \gamma_2(A) \leq 2^{\mathrm{Q}^{\mathrm{CC},*}(F)+1}.$$

Hence $\log \tilde{\gamma}_{2,\epsilon}(F) \leq \mathrm{Q}^{\mathrm{CC},*}(F) + 1$, as desired. □

Note that since the quantum communication complexity of F with shared entanglement is smaller than or equal to the randomized communication complexity of F with shared randomness, the measure $\log \tilde{\gamma}_2(F)$ also lower bounds $\mathrm{R}^{\mathrm{CC}}(F)$. Additionally, since $\tilde{\gamma}_{2,\epsilon}(F) = \gamma_2(A)$ for some A which ϵ-approximates F, and since $\gamma_2(A) \leq \|A\|_\infty \sqrt{\mathrm{rank}(A)} \leq \sqrt{\mathrm{rank}(A)} \leq \sqrt{\widetilde{\mathrm{rank}}_\epsilon(F)}$, we have

$$\log \tilde{\gamma}_{2,\epsilon}(F) \leq \frac{1}{2} \log \widetilde{\mathrm{rank}}_\epsilon(F).$$

Interestingly, there is also a relationship in the other direction, as shown by [LS09a].

Theorem 9.19. *For all (possibly partial) communication functions F with $\mathrm{Dom}(F) \subseteq \mathcal{X} \times \mathcal{Y}$ and all $\epsilon \in (0, 1/2)$, we have*

$$\log \widetilde{\mathrm{rank}}_\epsilon(F) = O(\log \tilde{\gamma}_{2,\epsilon}(F) + \log\log(|\mathcal{X}||\mathcal{Y}|) + \log(1/\epsilon) + \log(1/(1-2\epsilon))).$$

In particular, when ϵ is constant, we have

$$\log \widetilde{\mathrm{rank}}(F) = O(\log \tilde{\gamma}_2(F) + \log\log(|\mathcal{X}||\mathcal{Y}|)).$$

Proof. Let A be a matrix which ϵ-approximates F in the $\{+1,-1\}$ representation (that is, $|A| \leq 1$ and $A[x,y]$ is within 2ϵ of $F[x,y]$ when $F[x,y] \neq *$) and has $\gamma_2(A) = \tilde{\gamma}_{2,\epsilon}(F)$. We wish to construct a matrix B which pointwise approximates A and which has low rank. We first note that since $\mu(A) \leq 4K_G\gamma_2(A)$, we can write A as a linear combination of rectangles, $A = \sum_R w_R R$, with $\sum_R |w_R| \leq T$, where $T = 4K_G\tilde{\gamma}_{2,\epsilon}(F)$. Consider the probability distribution α over rectangles defined by giving R probability $|w_R|/T$. Observe that $\mathbb{E}_{R\sim\alpha}[T \operatorname{sign}(w_R) R] = \sum_R w_R R = A$; hence the expectation of $\operatorname{sign}(w_R)R$ with R sampled from α is entry-wise equal to A/T.

Consider a random matrix B_k formed by sampling k rectangles independently from α, adding up $\operatorname{sign}(w_R)R$ for these rectangles, and then multiplying the result by T/k. Note that $\mathbb{E}[B_k] = A$ for each value of k. For each entry (x,y), the expectation of $B_k[x,y]$ is $A[x,y]$, and we now bound the probability that $B_k[x,y]$ is too far from its expectation. We can write $B_k[x,y]$ as T/k multiplied by a sum of k random variables, one for each $\operatorname{sign}(w_R)R[x,y]$ for a rectangle R sampled from α. In fact, for each (x,y), the value of $\operatorname{sign}(w_R)R[x,y]$ is always in $\{-1,0,1\}$; grouping together these events, we see that $B_k[x,y]$ is T/k times a sum of k independently sampled random variables $Z_1, Z_2, \ldots, Z_k$,

where Z_i has probability p of being 1, probability q of being -1, and probability $1-p-q$ of being 0, with $p+q \leq 1$.

Hoeffding's inequality then says

$$\Pr[|B_k[x,y] - A[x,y]| > Tt/k] \leq 2e^{\frac{-kt^2}{2}}.$$

Hence for any $t > 0$, the probability that $B_k[x,y]$ is not within Tt/k of $A[x,y]$ is at most $2e^{\frac{-kt^2}{2}}$. By the union bound, the probability that for any (x,y) we have that $B_k[x,y]$ is not within Tt/k of $A[x,y]$ is at most $2nme^{\frac{-kt^2}{2}}$, where m and n are the dimensions of A. If we pick $k = \lfloor 2\ln(2nm)/t^2 \rfloor + 1$, then this probability is strictly smaller than 1, and hence there exists *some* choice of B_k which is within Tt/k of A at each entry. Note that assuming $nm \geq 2$ and $t \leq 1$, we have $k \leq 6\ln(nm)/t^2$ and $k \geq \ln(nm)/t^2$. Hence for any t, there is a matrix B which is within $Tt^3/\ln(nm)$ of A entrywise and which is a sum of at most $6\ln(nm)/t^2$ rectangles, and hence has rank at most $6\ln(nm)/t^2$. Pick t so that $Tt^3/\ln(nm) = \delta$, i.e. $t = (\delta \ln(nm)/T)^{1/3}$. Then there is a matrix B which is entrywise within δ of A and has rank at most $6\ln(nm)^{1/3}T^{2/3}/\delta^{2/3}$. Dividing B by $1+\delta$ gets it to have entries of absolute value at most 1; the maximum distance between an entry of B and an entry of A after this scaling is $1 - (1-\delta)/(1+\delta) =\leq 2\delta$.

This lets us conclude that

$$\widetilde{\text{rank}}_{\epsilon+2\delta}(F) \leq \tilde{\gamma}_{2,\epsilon}(F)^{2/3} \cdot 6\ln(nm)^{1/3}/\delta^{2/3},$$

so

$$\log \widetilde{\text{rank}}_{\epsilon+2\delta}(F) \leq (2/3)\log \tilde{\gamma}_{2,\epsilon}(F) + (1/3)\log\log(nm) + (2/3)\log(1/\delta) + 4.$$

We now pick $\delta = \min\{\epsilon, 1-2\epsilon\}/4$. Doing so allows us to amplify log rank with error $\epsilon + 2\delta$ to log rank with error ϵ with only a constant factor overhead (plus an additive constant). This gives

$$\log \widetilde{\text{rank}}_{\epsilon}(F) = O(\log \tilde{\gamma}_{2,\epsilon}(F) + \log\log(|\mathcal{X}||\mathcal{Y}|) + \log(1/\epsilon) + \log(1/(1-2\epsilon))). \qquad \square$$

This theorem is somewhat reminiscent of the result $\mathrm{R}^{\mathrm{CC,PRIV}}(F) = O(\mathrm{R}^{\mathrm{CC}}(F) + \log\log(|\mathcal{X}||\mathcal{Y}|))$ that we showed last week. In fact, since we know that $\log \widetilde{\text{rank}}(F)$ lower bounds $\mathrm{R}^{\mathrm{CC,PRIV}}(F)$ and the no-entanglement measure $\mathrm{Q}^{\mathrm{CC}}(F)$, and since $\log \tilde{\gamma}_2(F)$ lower bounds $\mathrm{R}^{\mathrm{CC}}(F)$ and the entanglement-assisted $\mathrm{Q}^{\mathrm{CC},*}(F)$, it is reasonable to interpret the log-approximate-gamma-2-norm measure as the "shared randomness/entanglement" version of approximate logrank, or alternatively we might say that approximate logrank is the private-coin version of the approximate log gamma 2 norm.

9.5.1 Duality for the approximate gamma 2 norm

We've seen that $\gamma_2(F)$ can be written as a semidefinite program. We can similarly write $\tilde{\gamma}_2(F)$ as a semidefinite program, so that we can take its dual and get a maximization program (useful for lower bounds). However, it will be even simpler to write $\tilde{\mu}(F)$ as a linear program and take its dual instead. Here by $\tilde{\mu}(F)$, we mean the minimum value of $\mu(A)$ for a matrix A which ϵ-approximates

F (where by default $\epsilon = 1/3$). This can be written

$$\begin{array}{rlrll} \tilde{\mu}_\epsilon(F) = & \min & \eta & & \\ & \text{s.t.} & \sum_R z_R & \leq \eta & \\ & & w_R & \leq z_R & \forall R \\ & & -w_R & \leq z_R & \forall R \\ & & \sum_R w_R R[x,y] & = A[x,y] & \forall (x,y) \in \mathcal{X} \times \mathcal{Y} \\ & & A[x,y] & \leq 1 & \forall (x,y) \in \mathcal{X} \times \mathcal{Y} \\ & & -A[x,y] & \leq 1 & \forall (x,y) \in \mathcal{X} \times \mathcal{Y} \\ & & F[x,y]A[x,y] & \geq 1 - 2\epsilon & \forall (x,y) \in \text{Dom}(F). \end{array}$$

The dual of this can be written

$$\begin{array}{rlrll} \tilde{\mu}_\epsilon(F) = & \max & (1-2\epsilon)\langle J, D\rangle - \langle J, C_1\rangle - \langle J, C_{-1}\rangle & & \\ & \text{s.t.} & a_1[R] + a_{-1}[R] & \leq 1 & \forall R \\ & & \langle R, B\rangle + a_1[R] - a_{-1}[R] & = 0 & \forall R \\ & & C_1 - C_{-1} - B + F \circ D & = 0 & \\ & & D[x,y] & = 0 & \forall (x,y) \notin \text{Dom}(F) \\ & & C_1, C_{-1}, D, a_1, a_{-1} & \geq 0. & \end{array}$$

If $C_1[x,y] > 0$ and $C_{-1}[x,y]$ ever occurs for the same (x,y), we could decrease both and improve the objective value; hence we can assume $C_1 \circ C_{-1} = 0$, and replace both variables with the single unconstrained matrix $C = C_1 - C_{-1}$. Similarly, if $a_1[R] > 0$ and $a_{-1}[R] > 0$, we can decrease both without changing anything, so we can assume that $a_1 \circ a_{-1} = 0$ and set $a = a_1 - a_{-1}$. This gives

$$\begin{array}{rlrll} \tilde{\mu}_\epsilon(F) = & \max & (1-2\epsilon)\langle J, D\rangle - \langle J, |C|\rangle & & \\ & \text{s.t.} & |a[R]| & \leq 1 & \forall R \\ & & \langle R, B\rangle + a[R] & = 0 & \forall R \\ & & C - B + F \circ D & = 0 & \\ & & D[x,y] & = 0 & \forall (x,y) \notin \text{Dom}(F) \\ & & D & \geq 0. & \end{array}$$

It is easy to see that the first two constraints are equivalent to $|\langle R, B\rangle| \leq 1$ for all R, which is the constraint $\|B\|_C \leq 1$. Moreover, the third constraint is simply $B = C + F \circ D$, so the first three together are equivalent to $\|C + F \circ D\|_C \leq 1$.

$$\begin{array}{rlrll} \tilde{\mu}_\epsilon(F) = & \max & (1-2\epsilon)\langle J, D\rangle - \langle J, |C|\rangle & & \\ & \text{s.t.} & \|C + F \circ D\|_C & \leq 1 & \\ & & D[x,y] & = 0 & \forall (x,y) \notin \text{Dom}(F) \\ & & D & \geq 0. & \end{array}$$

Finally, note that moving $C[x,y]$ closer to 0 while moving $(F \circ D)[x,y]$ in the opposite direction increases the objective value. This possible unless either $C[x,y] = 0$ or else $D[x,y] = 0$ and (either $F[x,y] = *$ or $F[x,y]C[x,y] < 0$). Split $C = H + D' \circ F$ where D' is 0 outside $\text{Dom}(F)$ and H is 0 inside $\text{Dom}(F)$. Then H is unconstrained. The matrix in the constraint is $H + D' \circ F + D \circ F$, and if we write $E = (D' + D) \circ F$, then D constitutes the positive entries of $E \circ F$ while D' constitutes the negative entries of $E \circ F$. Thus the constraint becomes $\|E + H\|_C \leq 1$, and the objective value becomes $\langle J, (1-2\epsilon)D - |C|\rangle = \langle J, D - 2\epsilon D - |C|\rangle = \langle F, E\rangle - \langle J, |H|\rangle - 2\epsilon\langle J, D\rangle$, so we are trying to maximize

$$\frac{\langle F, E\rangle - \langle J, |H|\rangle - 2\epsilon\langle J, D\rangle}{\|E + H\|_C}.$$

We can scale $E + H$ so that the absolute values of its entries sum to 1, and then interpret it as $\mu \circ G$, where G is a sign matrix and μ is a probability distribution. Let p be the probability under μ that F and G disagree, let q be the probability mass of μ outside $\mathrm{Dom}(F)$, and let $1 - p - q$ be the probability F and G agree. The we are trying to maximize

$$\frac{1 - 2p - 2q - 2\epsilon(1 - p - q)}{\mathrm{Disc}_\mu(G)} = \frac{(1 - \epsilon)(1 - 2p - 2q) - \epsilon}{\mathrm{Disc}_\mu(G)}$$

$$= \frac{\Pr_{(x,y)\sim\mu}[F[x, y] = G[x, y]] - \Pr_{(x,y)\sim\mu}[F[x, y] \neq G[x, y]] - \epsilon}{\mathrm{Disc}_\mu(G)}.$$

This lower bound technique, which is tight for $\tilde{\gamma}_2(F)$, says that we can lower bound the latter by finding a modified function G and a probability distribution μ over the inputs such that G has low discrepancy with respect to μ, and such that G has good correlation with F against μ. This is "generalized discrepancy", an analogue of which we've also seen in query complexity. To show that F is hard, we just need to show that under some distribution μ, F has constant correlation with a function G which has low discrepancy (i.e. low PP complexity), where both the correlation and the discrepancy are computed against the same distribution μ. The quantum communication complexity of F (even with shared entanglement) will be at least the logarithm of this bound.

References

[BW01] Harry Buhrman and Ronald de Wolf. "Communication complexity lower bounds by polynomials". In: *Proceedings 16th Annual IEEE Conference on Computational Complexity*. 2001 (p. 5).

[LS09a] Troy Lee and Adi Shraibman. "An approximation algorithm for approximation rank". In: *2009 24th Annual IEEE Conference on Computational Complexity*. 2009 (p. 16).

[LS09b] Nati Linial and Adi Shraibman. "Lower bounds in communication complexity based on factorization norms". In: *Random Structures & Algorithms* 34.3 (2009) (p. 6).

[LSŠ08] Troy Lee, Adi Shraibman, and Robert Špalek. "A direct product theorem for discrepancy". In: *2008 23rd Annual IEEE Conference on Computational Complexity*. 2008 (pp. 6, 9).